MAY 2 6 1981

NAS 1.60:1859

NASA Technical Paper 1859

ORIGINAL
COMPLETED

Computer Simulation of Random Variables and Vectors With Arbitrary Probability Distribution Laws

Victor M. Bogdan

MAY 1981

NASA

NASA Technical Paper 1859

Computer Simulation of Random Variables and Vectors With Arbitrary Probability Distribution Laws

Victor M. Bogdan
Lyndon B. Johnson Space Center
Houston, Texas

NASA
National Aeronautics
and Space Administration

**Scientific and Technical
Information Branch**

1981

BLANK PAGE

BLANK PAGE

CONTENTS

Section		Page
1.0	INTRODUCTION	1
2.0	COMPUTERIZATION OPERATOR	1
3.0	SIMULATION OF A SINGLE RANDOM VARIABLE	4
4.0	SIMULATION OF INDEPENDENT RANDOM VARIABLES	6
5.0	INTEGRAL PROPERTY OF THE COMPUTERIZATION OPERATOR	8
6.0	EXISTENCE OF TRANSITION PROBABILITIES	12
7.0	RESOLUTION OF BOREL PROBABILITIES	18
8.0	EXISTENCE OF CONDITIONAL DISTRIBUTIONS	23
9.0	SIMULATION THEOREM	26
10.0	EXAMPLES OF SIMULATION	29
11.0	EXPECTED VALUE OF A FUNCTION OF A PROCESS	35
12.0	INTEGRAL FORMULA FOR EXPECTATION	36
13.0	CONCLUSION	37
14.0	REFERENCES	38

1.0 INTRODUCTION

In many problems of quantum chemistry, nuclear and plasma physics, and economics, one may encounter a random process in which the variables are not independent and which can have a discrete, continuous, or mixed spectrum. In this paper, an algorithm is presented which allows a computer simulation of any such random process $x_1, \ldots, x_k$ of real random variables if their joint distribution is known. Examples are presented to illustrate the theory.

Mathematical tools used in developing the theory are based on Kolmogorov's fundamental paper on probability theory (ref. 9) and some results of Halmos (ref. 8), Neveu (ref. 10), and Bogdanowicz (refs. 2 through 6) on measure and integration theory.

Readers who are interested in applications only should concentrate on sections 1 through 4 and 8 through 13 and read the remaining sections as needed to understand the principles. Knowledge of the Lebesgue integral with respect to an abstract measure is essential to understanding the proofs. The use of Dirac's delta function is helpful in applications.

The theoretical results are formulated in terms of Borel functions; that is, functions measurable with respect to the smallest sigma ring containing all cubes. As established by Halmos (ref. 8), this class of functions coincides for R^k spaces with the class of Baire functions; that is, the smallest class which is closed under the sequential limit and contains all continuous functions. The importance of Baire functions in the general theory of random processes is presented in reference 1.

2.0 COMPUTERIZATION OPERATOR

Let F be the probability distribution of a real random variable x; i.e., $F(a) = P\{x < a\}$ for all $a \in R$. Such a function has the following properties:

a. F is nondecreasing on R.

b. F is left side continuous on R.

c. $F(-\infty) = 0$ and $F(\infty) = 1$.

These properties characterize distributions of real random variables according to Kolmogorov's theorem; i.e., if F has properties a, b, and c, then there exists a probability space and a random variable x over it such that F is its probability distribution.

If a distribution F is absolutely continuous in the Lebesgue sense, then there exists a Lebesgue summable function f on R such that $F(a) = \int_{-\infty}^{a} f(t)dt$ for all $a \in R$. Such a function f is called the density of the distribution F.

In many applications, one encounters distributions that do not have Lebesgue summable densities. For example, for $x \in R$, let

$$f(x) = p_1\delta(x - x_1) + p_2\delta(x - x_2) + \ldots + p_n\delta(x - x_n) \qquad (2\text{-}1)$$

where

$0 \leq p_i$

$p_1 + \ldots + p_n = 1$

$\delta(x)$ denotes Dirac's delta function; i.e., the formal density of the distribution h given by the formula

$h(a) = 0$ if $a \leq 0$

$h(a) = 1$ if $a > 0$

Then the distribution F corresponding to the function f is given by the formula

$$F(a) = p_1h(a - x_1) + \ldots + p_nh(a - x_n) \qquad (2\text{-}2)$$

for all $a \in R$. Such a distribution has jumps at the points $x_1, \ldots, x_n$ and is constant between them. Hence, F is not an absolutely continuous function. A random variable x corresponding to such a distribution has only discrete states $x_1, \ldots, x_n$.

However, in some applications, one encounters random variables with discrete and continuous states. For example, consider a random variable with a density

$$g(x) = p_1\delta(x - x_1) + p_2h(x)e^{-x} \qquad (2\text{-}3)$$

where

$0 \leq p_i$

$p_1 + p_2 = 1$

h is the distribution of δ

Such a density could appear in a steady state process x representing the energy level of a particle if the source emits particles having the specific energy level $x = x_1$ with probability p_1 and having other energy levels $x \neq x_1$ with joint probability p_2.

To simulate random variables with mixed states, it is convenient to introduce the computerization operator c mapping a distribution F into a function G over the open unit interval $(0,1)$. This function G is defined by the formula $G(u) = \inf\{x \in R: u < F(x)\}$ for all $u \in (0,1)$.

2.1 THEOREM

The computerization operator c is well defined for every function F that is a distribution; i.e., that has properties a, b, and c. If the variable u has a uniform distribution on the open interval $(0,1)$, then the variable $x = G(u)$ has a probability distribution equal to the function F.

Proof. Take any $u \in (0,1)$. It follows from property c that there are two points x_1 and x_2, such that $F(x_1) < u < F(x_2)$. This implies that the set $A(u) = \{x \in R: u < F(x)\}$ is nonempty. It follows from property a that the number x_1 is a lower bound of the set $A(u)$. From the axiom of continuity, it follows that the function $G(u) = \inf A(u)$ is well defined. To compute the distribution of the variable $x = G(u)$, consider the set $H(a) = \{u \in (0,1): G(u) < a\}$ for a fixed $a \in R$.

It follows from the property of infimum that $u \in H(a)$ if and only if there exists an $x \in R$ such that $x < a$ and $u < F(x)$. Thus, introducing the set $\{u \in (0,1): u < F(x)\} = (0,F(x))$, we get $H(a) = U_{x<a}(0,F(x))$. Since the function F is continuous from the left, the union of the intervals $(0,F(x))$ over all $x < a$ is equal to the interval $(0,F(a))$. Thus, $H(a) = (0,F(a))$ for all $a \in R$. Since the probability that a uniformly distributed variable on the interval $(0,1)$ falls into an interval I being a subinterval of $(0,1)$ is equal to the length of that interval, we get $P\{x < a\} = P(H(a)) = P((0,F(a))) = F(a)$ for all $a \in R$.

2.2 REMARK

For every distribution F, the function $G = c(F)$ has the following properties:

1. G is nondecreasing on $(0,1)$.

2. G is right side continuous on $(0,1)$.

3. If $u \in (F(x), F(x+))$, when F has a jump at x, then $G(u) = F(X+)$.

4. If F is strictly increasing and continuous on a closed interval $\langle c,d\rangle$, then $G(u) = x$ for $u \in \langle G(c), G(d)\rangle$ if and only if $u = F(x)$.

These properties of the function G allow one to find the graph of the function G from the graph of the distribution function F by the following steps:

1. Fill all vertical jumps in the graph of the function $u = F(x)$ with linear segments.

2. Treat the u-axis as the axis of the independent variable and the x-axis as that of the dependent variable.

3. At points u where there are several values of x such that $u = F(x)$, define G so that $G(u) = G(u+)$.

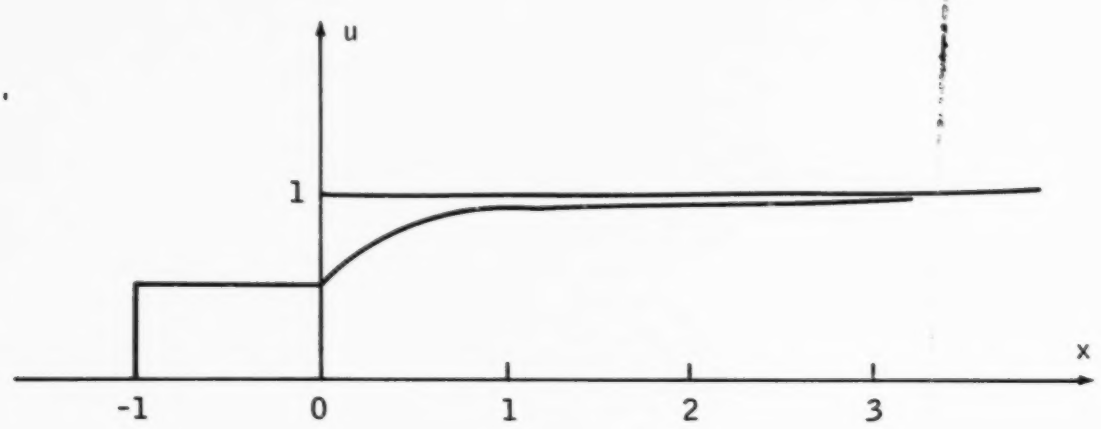
3.0 SIMULATION OF A SINGLE RANDOM VARIABLE

Since random number generators available on computers simulate a random variable u with uniform distribution on the interval $(0,1)$, one may simulate with good accuracy the distribution F of a random variable whose computerization G is a piecewise continuous function. Many distributions appearing in applied problems fall into this category.

3.1 EXAMPLE

Assume that each call to the Fortran function RAN(0) gives a different random number u in the interval $(0,1)$. Write a segment of a Fortran program to generate $N = 31$ random values of a variable with the density $g(x) = 0.5\delta(x + 1) + 0.5h(x)e^{-x}$, $x \in R$, where δ and h are defined as before.

Solution. Computing the distribution function from the density g by the formula $F(x) = \int_{-\infty}^{x} g(t)dt$ for all $x \in R$, we get $F(x) = 0.5h(x + 1) + 0.5h(x)(1 - e^{-x})$. The graph of the function F is given in the following diagram.



From this diagram, we get the formula for the computerization G of distribution F using remark 2.2. This yields

$$\left.\begin{array}{ll} G(u) = -1 & \text{if } 0 < u < 0.5 \\ G(u) = -\log(2 - 2u) & \text{if } 0.5 \leq u < 1 \end{array}\right\} \qquad (3\text{-}1)$$

Thus, the segment of the program to simulate the distribution F may look as follows:

```
      DIMENSION X(100)
      N = 31
      DO 5  I = 1, N
      X(I) = G(RAN(0))
    5 CONTINUE
      .
      .
      .

      FUNCTION G(U)
      IF (0.0.LT.U.AND.U.LT.0.5)  G = -1.
      IF (0.5.LE.U.AND.U.LT.1.)  G = -ALOG(2.-2.*U)
      RETURN
      END
```

The array X(I) for I = 1 to 31 will contain a random sample of a variable whose distribution is given by the function F.

3.2 EXAMPLE

Find the computerization G for a random variable whose density function is

$$\left.\begin{aligned} f(x) &= (1/2)\cos x \quad \text{if} \quad -\pi/2 < x < \pi/2 \\ f(x) &= 0 \qquad\qquad \text{if} \quad |x| \geq \pi/2 \end{aligned}\right\} \tag{3-2}$$

Solution. The distribution F of the function f is given by the formula

$$\left.\begin{aligned} F(x) &= (1/2)(\sin x + 1) \quad \text{if} \quad |x| \leq \pi/2 \\ F(x) &= 0 \qquad\qquad\qquad\quad \text{if} \quad x < -\pi/2 \\ F(x) &= 1 \qquad\qquad\qquad\quad \text{if} \quad x > \pi/2 \end{aligned}\right\} \tag{3-3}$$

Since the function F is continuous on the closed interval $\langle -\pi/2, \pi/2\rangle$ and maps it onto the interval $\langle 0,1\rangle$, we can use the equivalence of $x = G(u)$ with $u = F(x)$; i.e., $u = (1/2)(\sin x + 1)$. Solving this equation for x, we get $x = \arcsin(2u - 1)$. Thus, the computerization G is given by $G(u) = \arcsin(2u - 1)$ for $u \in (0,1)$.

4.0 SIMULATION OF INDEPENDENT RANDOM VARIABLES

In many applications, one has to investigate processes consisting of several random variables $f_1, f_2, \ldots, f_k$. Such variables are called independent if their joint distribution function F defined by

$$F(a_1, a_2, \ldots, a_k) = P\{f_1 < a_1, \ldots, f_k < a_k\} \tag{4-1}$$

for all $(a_1, \ldots, a_k) \in R^k$, can be represented in the form

$$F(a_1, a_2, \ldots, a_k) = F_1(a_1)F_2(a_2)\ldots F_k(a_k) \tag{4-2}$$

for all $(a_1, \ldots, a_k) \in R^k$.

Thus, to simulate such a process it is enough to find the computerizations $G_i = c(F_i)$ for $i = 1, \ldots, k$. Then, if $u_1, \ldots, u_k$ are k independent random variables each having uniform distribution on the interval $(0,1)$, the variables $x_1 = G_1(u_1)$, $x_2 = G_2(u_2)$, $\ldots$, $x_k = G_k(u_k)$ will have the distribution given by the function F.

4.1 PROBLEM

Let $p = (\lambda, \phi)$ be a random point on the sphere

$$S = \{(x, y, z): \; x^2 + y^2 + z^2 = 1\} \tag{4-3}$$

where λ and ϕ are its spherical coordinates. Simulate the uniform distribution on the sphere S.

Solution. The probability density f at the point $p \in S$ is given by the formula

$$f(p) = \frac{1}{4\pi} \quad \text{for} \quad p \in S \tag{4-4}$$

Using spherical coordinates, we can write a representation of the set S as

$$S = \left\{p = (\lambda, \phi): \quad 0 < \lambda < 2\pi, \quad -\frac{\pi}{2} < \phi < \frac{\pi}{2}\right\} \tag{4-5}$$

where we have neglected sets of measure zero; i.e., the poles and one meridian. The density function in these coordinates will have the form

$$f(\lambda, \phi) = f_1(\lambda) f_2(\phi), \quad (\lambda, \phi) \in S \tag{4-6}$$

where

$$f_1(\lambda) = 1/(2\pi) \qquad \text{for all} \quad \lambda \in (0, 2\pi)$$

$$f_2(\phi) = \frac{1}{2} \cos \phi \qquad \text{for all} \quad \phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

Thus, the joint distribution F of the variables λ, ϕ is given by

$$F(a_1, a_2) = \int_{-\infty}^{a_1} f_1(\lambda)d\lambda \int_{-\infty}^{a_2} f_2(\phi)d\phi = F_1(a_1)F_2(a_2) \tag{4-7}$$

for all $(a_1, a_2) \in R^2$, where F_1 is the distribution of λ and F_2 the distribution of ϕ. Since the joint distribution is a product of the two distributions, the random variables λ, ϕ are independent. Moreover,

$$\left.\begin{aligned} F_1(a_1) &= \frac{1}{2\pi} a_1 \quad \text{if} \quad 0 \leq a_1 \leq 2\pi \\ F_1(a_1) &= 0 \quad \text{if} \quad a_1 < 0 \\ F_1(a_1) &= 1 \quad \text{if} \quad a_1 > 2\pi \end{aligned}\right\} \tag{4-8}$$

The distribution F_2 was discussed in example 3.2. Thus, computerizations of the distributions F_1 and F_2 are given by

$$G_1(u_1) = 2\pi u_1 \tag{4-9}$$

$$G_2(u_2) = \text{arc sin } (2u_2 - 1) \tag{4-10}$$

Hence, the variables $\lambda = 2\pi u_1$ and $\phi = \text{arc sin } (2u_2 - 1)$, where u_1, u_2 are independent uniform random variables on the interval $(0,1)$, will simulate a uniform distribution of points on the sphere S.

5.0 INTEGRAL PROPERTY OF THE COMPUTERIZATION OPERATOR

Let F be the distribution of a real random variable, i.e., F satisfies conditions a, b, c of section 2. Let $G = c(F)$ be the computerization of the distribution F. Denote by V the prering (see ref. 2) consisting of all intervals I of the form $<a,b)$, $(-\infty,a)$, $<a,\infty)$, where a and b are real numbers. Define a set function v on V by the formula

$$\left.\begin{aligned} v<a,b) &= F(b) - F(a) \\ v(-\infty,b) &= F(b) - F(-\infty) = F(b) \\ v<a,\infty) &= F(\infty) - F(a) = 1 - F(a) \end{aligned}\right\} \tag{5-1}$$

We can prove that the set function v is countably additive on V and thus forms a volume in the sense of Bogdanowicz (ref. 2). Following the development of paper 2, denote by $S(V, R)$ the collection of simple functions; i.e., functions of the form

$$s(x) = r_1 c_{A_1}(x) + \ldots + r_k c_{A_k}(x) \quad \text{for all} \quad x \in R \tag{5-2}$$

where $A_1, \ldots, A_k$ are disjoint sets from the prering V; $r_1, \ldots, r_k$ are real numbers; and c_A denotes the characteristic function of the set A. The set of simple functions is linear and the following functionals are well defined on it:

$$\int s dv = r_1 v(A_1) + \ldots + r_k v(A_k) \tag{5-3}$$

$$||s|| = |r_1| v(A_1) + \ldots + |r_k| v(A_k) \tag{5-4}$$

The first functional is linear and the second forms a seminorm on $S(V, R)$. Moreover, $|\int s dv| \leq ||s||$ for all simple functions.

Denote by N the collection of all sets A of R such that for every $\varepsilon > 0$ there exists a countable family $A_t \in V$ $(t \in T)$ such that the set A is contained in the union $U_T A_t$ and $\Sigma_T v(A_t) < \varepsilon$. Sets of this collection N will be called v-null sets.

A sequence $s_n \in S(V, R)$ is called basic if there exists a sequence k_n of simple functions and a constant M such that $s_n = k_1 + k_2 + \ldots + k_n$, $||k_n|| \leq M4^{-n}$ for all n. Denote by $L(v, R)$ the set of all functions f for which there exists a basic sequence s_n and a null set $A \in N$ such that the sequence of values $s_n(x)$ converges to the value $f(x)$ if $x \notin A$.

Define $||f|| = \lim ||s_n||$, $\int f dv = \lim \int s_n dv$. According to paper 2, these are well-defined functionals on $L(v, R)$; and the space $L(v, R)$ coincides with the space of Lebesgue summable functions with respect to the Lebesgue measure p, which is the smallest complete measure extending the volume v (see ref. 5). Moreover, the two integrals, $\int f dv$ and $\int f dp$, coincide.

In the sequel, we shall write $\int_a^b f(x) dF(x)$ to denote the integral $\int c_{<a,b)} f dv$.

Notice that the classical Lebesgue integral is generated by the function $g(u) = u$ for all $u \in R$, which corresponds to the volume $v(<a,b)) = b - a$ on the prering W of all bounded right side open intervals.

We shall say that a function f is v-summable on a set A, or equivalently that the integral $\int_A f dv$ exists, if and only if $c_A f \in L(v, R)$.

5.1 THEOREM

Let F be a probability distribution over R and G its computerization. Then, if the right-hand integral in the following formula exists in Lebesgue's sense, so does the other and they are equal

$$\int_R f(x)dF(x) = \int_0^1 f(G(u))du \tag{5-5}$$

5.2 REMARK

The theorem is valid for the Riemann-Stieltjes integral when the function F is continuous and invertible. Notice that in theorem 5.1 each function, F and G, may have an infinite number of discontinuities and neither has to be invertible.

5.3 REMARK

Let v be the volume generated by the distribution F and μ the classical Lebesgue measure over the interval $(0,1)$. The above theorem is equivalent to the following. The map $f \to foG$ imbeds isometrically the Lebesgue space $L(v, R)$ into the space $L(\mu, R)$.

Proof of the theorem. Let $f = c_{(-\infty,a)}$. Then $foG = c_{(0, F(0))}$. Indeed $foG(u) = 1$ if and only if $G(u) \in (-\infty,a)$; i.e., $G(u) < a$, which is equivalent, as proved in section 2, to $u \in (0, F(a))$; i.e., to $c(0, F(a))(u) = 1$. Thus, the characteristic function of an interval $<a,b)$ is mapped into the characteristic function of the interval $<F(a), F(b))$. Indeed,

$$\begin{aligned} c_{<a,b)}oG &= (c_{(-\infty,b)} - c_{(-\infty,a)})oG = c_{(-\infty,a)}oG - c_{(-\infty,b)}oG \\ &= c_{(0, F(b))} - c_{(0, F(a))} = c_{<F(a), F(b))} \end{aligned} \tag{5-6}$$

Similarly, the characteristic function of $(-\infty,b)$ is mapped into the characteristic function of $(0, F(b))$ and the characteristic function of $<a,\infty)$ into the characteristic function of $<F(a), 1)$. By the definition of the volume v on the prering V and of the Lebesgue measure μ, we get

$v(I) = \mu(G^{-1}(I))$ for every $I \in V$, since $c_I oG = c_A$ is equivalent to $A = G^{-1}(I)$. These observations yield the equality

$$\int |s| dv = \int |soG| d\mu \quad \text{for all} \quad s \in S(V, R) \tag{5-7}$$

Let W be the collection of all left side closed subintervals of the interval $(0,1)$. It follows from the definition of a v-null set that if A is a v-null set then the μ measure of the set $B = G^{-1}(A)$ is zero. Notice that if $s_n \in S(V, R)$ is a basic sequence convergent for all $x \notin A$ to the function f then the sequence $s_n oG$ belongs to the set $S(W, R)$ and converges for all points $u \notin B$ to the function foG. Thus, from the Bogdanowicz definition of the spaces $L(v, R)$ and $L(\mu, R)$, we get

$$\int |f| dv = \lim \int |s_n| dv = \lim \int |s_n oG| d\mu = \int |foG| d\mu \tag{5-8}$$

for all functions $f \in L(v, R)$. This proves the theorem.

5.4 COROLLARY

For every bounded Borel function f and every $a \in R$ the following equality holds:

$$\int_{-\infty}^{a} f(x) dF(x) = \int_{0}^{F(a)} f(G(u)) du \tag{5-9}$$

Proof. It follows from the properties of v-measurable functions (ref. 3) that every Borel measurable function is v-measurable. Since $c_{(-\infty,a)}$ is a Borel function, the product $g = c_{(-\infty,a)} f$ also is a Borel function. Being v-measurable and bounded by the simple function $Mc_R = Mc_{(-\infty,a)} + Mc_{<a,\infty)}$ for some M, the function g is v-summable. Since $goG = c_{(0,\ F(a))} foG$, we get from theorem 5.1 the equality

$$\int_{-\infty}^{a} f(x) dF(x) = \int_{0}^{F(a)} f(G(u)) du \tag{5-9}$$

6.0 EXISTENCE OF TRANSITION PROBABILITIES

Let $f_1, f_2, \ldots, f_k$ be real random variables over a probability space. We shall prove that the following conditional distribution

$$P\{f_1 < a_1 | f_2 = a_2, f_3 = a_3, \ldots, f_k = a_k\} \tag{6-1}$$

can be well defined as a Borel function of the vector $a = (a_1, a_2, \ldots, a_k)$ over the space R^k.

Let F_m denote the joint distribution of the variables $f_1, f_2, \ldots, f_m$. Each distribution function F obtained in this way is nondecreasing with respect to the relation $a < b$ on R^m defined to mean $a_i < b_i$ for all $i = 1, \ldots, m$. This means that if $a < b$ then $F(a) \leq F(b)$. Moreover, the distribution function is continuous with respect to increasing convergence; i.e., the condition $a_j^n \uparrow a_j$ for $j = 1, \ldots, m$ implies that $F(a_1^n, a_2^n, \ldots, a_m^n) \to F(a_1, a_2, \ldots, a_m)$.

Finally, the function F is normalized; i.e., $F(-\infty, \ldots, -\infty) = 0$, $F(\infty, \ldots, \infty) = 1$.

According to Kolmogorov's theorem, these properties characterize a joint distribution function F; i.e., for every such function there exists a unique Borel probability measure P over R^m such that

$$F(a_1, \ldots, a_m) = P\{e_1 < a_1, \ldots, e_m < a_m\} \tag{6-2}$$

for all $(a_1, \ldots, a_m) \in R^m$, where e_j are projection functions defined by

$$e_j(a_1, \ldots, a_m) = a_j \quad \text{for all } (a_1, \ldots, a_m) \in R^m$$

To make the presentation more general, it will be convenient to introduce the following notation. If M and K are subsets of the set $\{1, \ldots, k\}$ and K is a proper subject of M, then we will write $K < M$. Subsets of this form will be called indexes. The symbol $|K|$ will denote the number of elements of the set K.

We will denote by R^K the space of all vectors $x = (x_t)_{t \in K}$, where $x_t \in R$ denotes the component of the vector x with index t. If M and K are two disjoint index sets and their union is $S = M \cup K$, then the space R^S can be identified with the product space $R^M \times R^K$ and every vector $x \in R^S$ can be written in the form $x = (x_K, x_M)$, where $x_K \in R^K$ and $x_M \in R^M$.

If $a \in R^T$, we shall denote by $I(a)$ the Cartesian product $\times_{t \in T}(-\infty, a_t)$, where $a = (a_t)_{t \in T}$. Sets of this form will be called in the sequel basic cones.

Now if F_T is a probability distribution on R^T and p_T is the corresponding Borel probability over R^T obtained from Kolmogorov's theorem (i.e., $F_T(a) = p_T(I(a))$ for all $a \in R^T$), for disjoint decomposition of the index set T into nonempty sets S and U the Borel probability p_S is well defined by the formula

$$p_S(A) = p_T(R^U \times A) \tag{6-3}$$

for all Borel subsets A of the space R^S.

This probability in turn generates a probability distribution F_S. A function p_U^S will be called a <u>transition probability</u> from probability p_S to probability p_T if its value $p_U^S(A, x)$ is defined for every Borel set A being a subset of the space R^U and every $x \in R^S$. Moreover, the value $p_U^S(A, x)$ as a function of set A is a probability measure for every $x \in R^S$, and as a function of point x is Borel for every fixed Borel set A. Furthermore,

$$p_T(A \times B) = \int_B p_U^S(A, x) p_S(dx) \tag{6-4}$$

for every Borel subset A of R^U and every Borel subset B of R^S (see ref. 10, p. 73).

6.1 THEOREM

For every Borel probability p_T over R^T, where T is finite, and every generated Borel probability p_S, where S is a subset of T and the difference set $U = \{x \in T: x \notin S\}$ is nonempty, there exists a transition probability p_U^S from the measure p_S to the measure p_T.

Proof. For every fixed Borel set A contained in R^U, let q_A denote the measure defined by the formula

$$q_A(B) = p_T(A \times B) \tag{6-5}$$

for all Borel sets B contained in R^S.

Since

$$q_A(B) \leq p_T(R^U \times B) = p_S(B) \tag{6-6}$$

for all Borel sets B in the space R^S, we get from the Radon-Nikodym theorem (see refs. 6 and 10) that there exists a Borel function f_A summable with respect to the measure p_S. After a modification on a Borel set of p_S-measure zero, we get

$$0 \leq f_A(x) \leq 1 \quad \text{for all} \quad x \in R^S \tag{6-7}$$

and

$$q_A(B) = \int_B f_A(x) p_S(dx) \tag{6-8}$$

for all Borel sets B contained in the space R^S.

If $b \in R^U$, let $I(b)$ denote the Cartesian product $\times_{t \in U}(-\infty, b_t)$. Every such set $I(b)$ is a Borel set. A vector $b \in R^U$ will be called rational if all its components b_t are rational. We shall write $a < b$ for two such vectors if and only if $a_t < b_t$ for all $t \in U$.

Let h be a function given by the formula

$$h(a, x) = \sup\{0, f_{I(b)}(x): \ b < a; \ b \text{ is rational}\} \tag{6-9}$$

for all $a \in R^U$ and all $x \in R^S$. The set following the supremum operation is nonempty since it contains zero and is bounded. Thus, the function h by the axiom of continuity is well defined and is a Borel function in variable x.

Since both the measure and the integral are continuous under increasing sequential convergence, we get from (6-8) and (6-5) the relation

$$p_T(I(a) \times B) = \int_B h(a, x) p_S(dx) \tag{6-10}$$

for all $a \in R^U$ and all Borel sets B of R^U. Moreover, from the definition of function h we get that, for every fixed x, it is nondecreasing; i.e., if for two vectors a, $c \in R^U$ we have $a < c$, then $h(a, x) \leq h(c, x)$. It is also left side continuous; i.e., if $a^n < a$ for all n and the vectors a^n converge to the vector a, then the sequence of values $h(a^n, x)$ converges to the value $h(a, x)$.

Now from relation (6-10), using the monotone convergence theorem, we get the following relations:

$$p_S(B) = \int_B h(\infty, x) p_S(dx) \tag{6-11}$$

$$0 = \int_B h(-\infty, x) p_S(dx) \tag{6-12}$$

for all Borel sets B in R^S, where $h(\infty, x)$ denotes the limit in the variable a all of whose coordinates tend to ∞. The value $h(-\infty, x)$ is understood similarly. Since

$$h(\infty, x) = \lim h(a^n, x) \quad \text{for all } x \in R^S \tag{6-13}$$

where a^n is an increasing sequence of vectors such that each component a_t^n tends to infinity, we get that $h(\infty, x)$ as a function of x is Borel measurable on the space R^S.

Thus, from Radon-Nikodym theorem there exists a set C of p_S-measure zero such that $h(\infty, x) = 1$ and $h(-\infty, x) = 0$ if $x \notin C$.

Modifying h on this set by putting $h(a, x) = g(a)$ for all $x \in C$, where g is any probability distribution on R^U, we get that the value $h(a, x)$ as a function of x is a Borel function for every a and $h(a, x)$ as a function of a is a probability distribution for every fixed x. Thus, by Kolmogorov's theorem h generates a unique probability measure $p_U^S(A, x)$ defined on all Borel sets A of R^U for every fixed point $x \in R^S$.

Let us prove that for every fixed Borel set A of R^U the following two properties hold:

A. The function $p_U^S(A, x)$ as a function of x is Borel measurable over the space R^S.

B. For every Borel set B of R^S, we have $p_T(A \times B) = \int_B p_U^S(A, x)\, p_S(dx)$.

To this end, denote by M the collection of all Borel sets A of R^U for which properties A and B hold.

Observe that the sets $I(a)$ belong to M. Since M is closed under disjoint finite union, it is also closed under proper differences; i.e., if A_1 is a subset of A_2 and both sets A_1, A_2 are in M, then also the difference set $A = \{x \in A_1: x \notin A_2\}$ is in M. Thus, if V denotes the prering consisting of all intervals of the form $(-\infty,a)$, and $<b,a)$, where a and b are real numbers, and V^U denotes the prering consisting of all Cartesian products of the form $A = \times_{t\in U} A_t$, where $A_t \in V$ for every $t \in U$, we can prove by induction with respect to the number of bounded intervals A_t appearing in the representation of the set A that the prering $W = V^U$ is contained in the collection M. Finite disjoint unions of sets from the prering W form the smallest ring containing W.

Observe that M is closed under monotone convergence of sets. According to a theorem of Halmos (ref. 8), this implies that M contains the smallest sigma ring generated by W. We can prove that this sigma ring coincides with the sigma ring of all Borel sets of the space R^U. This concludes the proof of the theorem.

6.2 THEOREM

Let T, S, U be as before and q_U^S, p_U^S be two transition probabilities from probability p_S to probability p_T.

There exists a Borel set C of p_S-measure zero such that

$$p_U^S(A, x) = q_U^S(A, x) \tag{6-14}$$

for all Borel sets A of R^U and all $x \notin C$.

Proof. It follows from the definition of a transition probability and from the Radon-Nikodym theorem that for every Borel set A of R^U there exists a Borel set $C(A)$ of p_S-measure zero such that

$$p_U^S(A, x) = q_U^S(A, x) \tag{6-15}$$

for all $x \notin C(A)$. Let D denote the set of all rational points b of the space R^U and let C be the union of the sets $C(I(b))$ over all $b \in D$.

Clearly, C is a Borel set of p_S-measure zero.

Denote by M the collection of all Borel sets A of R^U such that $p_U^S(A, x) = q_U^S(A, x)$ if $x \in C$. This collection contains basic cones $I(b)$, where b is a rational vector. It follows from the monotone continuity of a measure that M contains every set $I(a)$ for any $a \in R^U$. The rest of the argument is the same as in theorem 6.1. This concludes the proof.

Let a function F_U^S be defined on the product $R^U \times R^S$ such that for every fixed $x \in R^S$ the value $F_U^S(a, x)$ considered as a function of the variable a is a probability distribution and for every fixed $a \in R^U$ $F_U^S(a, x)$ considered as a function of x is a Borel function. Such a function F_U^S will be called in the sequel a transition distribution.

6.3 THEOREM

There is a one-to-one correspondence between transition probabilities q_U^S and transition distributions F_U^S. This correspondence is given by the relation

$$q_U^S(I(a), x) = F_U^S(a, x) \tag{6-16}$$

for all $a \in R^U$ and $x \in R^S$.

Proof. It is clear that every transition probability generates a transition distribution by means of formula (6-16). To prove that every transition distribution generates a unique transition probability, take an arbitrary fixed point $x \in R^S$ and denote by $q_U^S(A, x)$ the value of the probability measure defined for a Borel set A of R_U^S and generated from the probability distribution $F_U^S(a, x)$ by means of Kolmogorov's construction.

To prove that q_U^S is a transition probability, it is sufficient to prove that for every fixed Borel set A the value $q_U^S(A, x)$ as a function of $x \in R^S$ is Borel measurable. To this end, denote by M the collection of

all Borel sets A having that property. Notice that all basic cones I(a) belong to M. Notice that M is closed under disjoint finite union and under the monotone convergence of sets. Thus, as in the proof of theorem 6.1, we conclude that M coincides with the sigma ring of all Borel sets of R^U. This concludes the proof of the theorem.

7.0 RESOLUTION OF BOREL PROBABILITIES

Let T be a finite index set and U, S its disjoint decomposition into nonempty sets. Let p_T be a Borel probability over R^T and p_S the corresponding probability generated over R^S.

The theorems of the previous section show that the measure p_T generates almost unique representation of p_T by means of q_U^S and p_S through the formula

$$p_T(A \times B) = \int_B q_U^S(A, x)p_S(dx) \tag{7-1}$$

for all Borel sets A of R^U and B of R^S. Conversely, any pair q_U^S, p_S consisting of a transition probability and a probability on Borel sets generates a unique Borel probability over the space R^T (ref. 10, p. 74).

It follows from Kolmogorov's theorem that condition (7-1) is equivalent to

$$p_T(I(a) \times I(b)) = \int_{I(b)} q_U^S(I(a), x)p_S(dx) \tag{7-2}$$

for all $a \in R^U$ and all $b \in R^S$.

The necessity of condition (7-2) is obvious. To prove its sufficiency, fix the point $a \in R^U$ and consider two measures

$$r_1(B) = p_T(I(a) \times B) \tag{7-3}$$

$$r_2(B) = \int_B q_U^S(I(a), x)p_S(dx) \tag{7-4}$$

for all Borel sets B of R^S. Since these measures coincide for every basic cone according to relation (7-2), they must coincide for $B = R^S$. If $r_1(R^S) = 0$, then both measures are identically zero on all Borel sets B of R^S. If $r = r_1(R^S) > 0$, then dividing the measures r_1 and r_2 by the value r we get two probability measures that coincide on all basic cones $I(b)$. Thus, these two probability measures have the same distribution function. By Kolmogorov's uniqueness theorem, this implies

$$r_1(B)/r = r_2(B)/r \tag{7-5}$$

for all Borel sets B of R^U. Hence,

$$p_T(I(a) \times B) = \int_B q_U^S(I(a), x)p_S(dx) \tag{7-6}$$

for all Borel sets B of R^U and all basic cones $I(a)$, $a \in R^U$.

Holding the Borel set B fixed, by a similar argument to the preceding one, we get relation (7-1) for all Borel sets A of R^U and all Borel sets B of R^S. Now let us introduce the following notations for points in the spaces R^T and R^S. If $a \in R^T$, then by a_S, where S is a subset of T, we shall denote a point in the space R^S such that its component having index $t \in S$ coincides with the component a_t of a. Thus, we have $a = a_T$. When $S = \{t\}$, we shall write a_t instead of $a\{t\}$. If U and S are disjoint nonempty sets whose union is the set T, then the vector a_T can be identified with the pair (a_U, a_S). Notice also that $I(a_T) = I(a_U) \times I(a_S)$. Using this convention, we can write relation (7-2) in the equivalent form

$$p_T(I(a_T)) = \int_{I(a_S)} q_U^S(I(a_U), x_S)p_S(dx_S) \tag{7-7}$$

for all $a \in R^T$.

It is convenient to introduce a shorthand notation and convention similar to Einstein's convention in tensorial calculus. Namely, the relation given by formula (7-1) we shall write as

$$p_T = q_U^S p_S \tag{7-8}$$

This will mean that whenever in such a formula a superscript index set S coincides with the subscript index set, we have an integration over a Borel set with respect to the variable x_S.

Notice that the operator E mapping a pair (q_U^S, p_S) into the element $q_U^S p_S$ preserves convex combinations in both variables q_U^S and p_S. Thus, it is natural to extend it by homogeneity and linearity onto the space B_U^S of all Borel transition measures (since we will not use this space in the sequel, we leave it to the reader to give precise definition to this space) and onto the space B_S of all finite Borel measures over the space R^S.

Now let us extend the definition of the transition measure p_U^S to include the case when either S or U is empty. It is clear that when $U = 0$ the transition measure p_U^S is only a point function and the relation $q_S = p_U^S p_S$ means that p_U^S is the Radon-Nikodym derivative of the measure q_S with respect to the measure p_S. When $S = 0$, the transition measure p_U^S does not depend on the point and thus is a probability measure. Thus, we assume $p_U^0 = p_U$.

Finally, let $p_0 = 1$. Then the relation $q_T = p_U^S p_S$ defines uniquely the element q_T for any disjoint decomposition S, U of the index set T. If the index set S consists of a single point t, we shall write q_U^t instead of $q_U^{\{t\}}$ and similarly for the set U. Let $T = \{1, 2, \ldots, n\}$. The notation $T(j) = \{k: k < j\}$ will be used for $j = 1, 2, \ldots, n+1$. Notice that $T(1)$ denotes the empty set. A sequence of transition probabilities

$$q_j^{T(j)} \quad \text{for} \quad j = 1, 2, \ldots, n$$

will be called a resolution of the probability p_T if and only if

$$p_{T(j+1)} = q_j^{T(j)} p_{T(j)} \quad \text{for} \quad j = 1, 2, \ldots, n \tag{7-9}$$

It follows from the theorems in the previous section that such transition probabilities exist and are almost unique; i.e., $q_j^{T(j)}$ is unique up to a set of $p_{T(j)}$-measure zero.

Notice that the value of the transition probability $q_j^{T(j)}(A, x)$ gives precise meaning to the conditional probability

$$p_T\{f_j \in A | f_k = a_k \text{ for } k = 1, 2, \ldots, j-1\} \tag{7-10}$$

if the probability measure p_T is generated by joint distribution of the functions f_t $(t \in T)$. Since by theorem 6.1 there exists a transition probability $p_{S(j)}^1$ from the measure p_1 to measure $p_{T(j)}$, we can write

$$p_{T(j)} = p_{S(j)}^1 p_1 \tag{7-11}$$

for $j = 1, \ldots, n$, where $S(j) = \{2, 3, \ldots, j-1\}$.

7.1 THEOREM

If the sequence $q_j^{T(j)}(j = 1, \ldots, n)$ represents a resolution of the probability p_T, then there exists a Borel set C of p_1-measure zero such that for every fixed value $x_1 \notin C$ the sequence

$$q_j^{T(j)}(A, x_{S(j)}, x_1)$$

for $j = 2, \ldots, n$, as a function of A and the remaining variables x_t $(t \neq 1)$, represents a resolution of the Borel probability p_S defined by the formula

$$p_S(A) = p_S^1(A, x_1) \tag{7-12}$$

for all Borel sets A of R^S, where $S = \{2, 3, \ldots, n\}$.

Proof. From the definition of the resolution of the probability p_T, we have

$$p_{T(j+1)} = q_j^{T(j)} p_{T(j)} \qquad \text{for } j = 1, 2, \ldots, n \tag{7-13}$$

It follows from the reduction formula for integrals with respect to a probability measure generated by a transition probability (see Neveu, ref. 10, p. 74) that

$$\begin{aligned} & p_{T(j+1)}\,(I(a_{T(j+1)})) \\ & \quad = \int_{I(a_{T(j)})} q_j^{T(j)}(I(a_j), x_{T(j)})\, p_{T(j)}(dx_{T(j)}) \\ & \quad = \int_{I(a_1)} \Big(\int_{I(a_{S(j)})} q_j^{T(j)}(I(aj), x_{S(j)}, x_1) \\ & \qquad p^1_{S(j)}(dx_{S(j)}, x_1)\Big) p_1(dx_1) \\ & \quad = \int_{I(a_1)} r^1_{S(j+1)}(I(a_{S(j+1)}), x_1)\, p_1(dx_1) \end{aligned} \tag{7-14}$$

where $r^1_{S(j+1)}$ denotes the transition probability satisfying the condition

$$\begin{aligned} & r^1_{S(j+1)}(I(a_{S(j+1)}), x_1) \\ & \quad = \int_{I(a_{S(j)})} q_j^{T(j)}(I(a_j), x_{S(j)}, x_1)\, p^1_{S(j)}(dx_{S(j)}, x_1) \end{aligned} \tag{7-15}$$

This expression as a function of x_1 (according to ref. 10, p. 74) represents a Borel measurable function and as a function of the variable $a_{S(j+1)}$ represents a probability distribution for every fixed x_1. Hence, by theorem 6.3, it determines a unique transition probability $r^1_{S(j+1)}$. In this way from formula (7-14) we obtain the representation

$$p_{T(j+1)} = r^1_{S(j+1)}\, p_1 \tag{7-16}$$

It follows from the uniqueness of transition probability from $p_{T(j+1)}$ to p_1, theorem 6.2, that for some Borel set C of p_1-measure zero we have

$$p^1_{S(j+1)}(I(a_{S(j+1)}),\ x_1) = r^1_{S(j+1)}(I(a_{S(j+1)}),\ x_1) \tag{7-17}$$

for all $a_{S(j+1)} \in R^{S(j+1)}$ and all $x_1 \notin C$ and $j = 2, \ldots, n$.

Equalities (7-15) and (7-17) yield

$$p^1_{S(j+1)}(I(a_{S(j+1)}),\ x_1)$$

$$= \int_{I(a_{S(j)})} q_j^{T(j)}(I(a_j),\ x_{S(j)},\ x_1)\ p^1_{S(j)}(dx_{S(j)},\ x_1) \tag{7-18}$$

For a fixed $x_1 \notin C$, define the sequence of probabilities by

$$p_{S(j)}(A) = p^1_{S(j)}(A,\ x_1) \tag{7-19}$$

for all Borel sets A of $R^{S(j)}$ and $j = 2, \ldots, n$, and define a sequence of transition probabilities by the formula

$$Q_j^{S(j)}(A,\ x_{S(j)}) = q_j^{T(j)}(A,\ x_{S(j)},\ x_1) \tag{7-20}$$

for all Borel sets A and $j = 2, \ldots, n$.

We can prove from relation (7-18) that

$$p_{S(j+1)}(A \times R) = p_{S(j)}(A) \tag{7-21}$$

for all Borel sets A of $R^{S(j)}$ and $j = 2, 3, \ldots, n$. Relations (7-18), (7-19), (7-20), and (7-21) prove that the sequence $Q_j^{S(j)}$ represents a resolution of the probability $p_{S(n+1)} = p_S$. This completes the proof of the theorem.

8.0 EXISTENCE OF CONDITIONAL DISTRIBUTIONS

Let $q_j^{T(j)}$ $(j = 1, \ldots, n)$ be a resolution of the probability p_T as defined before. Let q_t^S denote one of the transition probabilities from this sequence. Define the function F_t^S by the formula

$$F_t^S(a_t, a_S) = q_t^S(I(a_t), a_S) \tag{8-1}$$

for all $a \in R^U$, where U is the union of the sets S and $\{t\}$.

These functions will be called the conditional distributions and the sequence $F_j^{T(j)}$ $(j = 1, \ldots, n)$ a resolution of the distribution F_T.

The conditional distribution F_t^S gives a precise meaning to the conditional probability

$$p_U\{f_t < a_t | f_j = a_j \text{ for } j \in S\} = F_t^S(a_t, a_S) \tag{8-2}$$

for all $a \in R^U$, where p_U is the probability generated by the process f_t $(t \in T)$ over R^U.

8.1 THEOREM

Every conditional distribution F_t^S, where $t \notin S$, is a Borel function over the space R^U.

Proof. Let $Z = \{b_1, b_2, \ldots\}$ be the set of all rational points and $Z_n = \{b_1, \ldots, b_n\}$. For every fixed $b \in Z$, the value $F_t^S(b, a_S)$ as a function of a_S on R^S is Borel, as follows from its definition. Define the function H_n by the formula

$$H_n(a_t, a_S) = \sup\{0, F_t^S(b, a_S): \ b \in Z_n, \ b < a_t\} \tag{8-3}$$

for all $a_t \in R$ and $a_S \in R^S$.

These functions are well defined and the domain of the variable $a_t \in R$ can be split into a finite number of disjoint intervals I_j $(j = 0, 1, \ldots, n)$ by means of the points of the set Z_n so that for each of these intervals the function $H_n(a_t, a_S)$ does not depend on a_t and is Borel in the variable $a_S \in R^S$. This follows from the fact that Borel functions are closed under the finite supremum operation, and that on the interval I_j only a finite number of elements of the set Z_n is smaller than a_t. Thus, every function H_n can be represented in the form

$$H_n(a_t, a_S) = \Sigma_j \, c_{I_j}(a_t) \, G_{jn}(a_S) \tag{8-4}$$

where G_{jn} are Borel functions on R^S and c_I is the characteristic function of the interval I. Since the characteristic function of an interval is a Borel function and the composition of a Borel function with a continuous function is a Borel function, we may consider the value $c_I(a_t) = c_I \circ e_t(a_U)$ as a function of a_U, where e_t is the projection function defined by $e_t(a_U) = a_t$ for all $a_U \in R^U$, as a Borel function over R^U. Similarly, we may consider the function $G_{jn}(a_S) = G_{jn} \circ e_S(a_U)$ as a Borel function over R^U.

Since Borel functions are closed under multiplication and addition, the functions H_n are Borel over R^U. Now notice that

$$F_t^S(a_t, a_S) = \lim H_n(a_t, a_S) \quad \text{for all } a \in R^U \tag{8-5}$$

Thus, the function F_t^S is a Borel function over the space R^U. This completes the proof.

Notice that every probability distribution F_S on R^S generates a unique volume v on the product prering V^S consisting of all sets of the form $A = \times_{t \in S} A_t$, where $A_t \in V$ for all $t \in S$, and V consists of all intervals of the form $(-\infty, a)$ and $\langle a, b)$ (see ref. 5). The Bogdanowicz integral with respect to the volume v coincides with the Lebesgue integral generated by the probability measure p_S.

Thus, the integral $\int f(x_S)v(dx_S)$ is uniquely determined if the distribution function F_S is known. By an integral with respect to the distribution F_S, we shall understand

$$\int f(x_S)F_S(dx_S) = \int f(x)v(dx) \tag{8-6}$$

Consequently, each function $F_j^{T(j)}$ from the sequence representing the resolution of the distribution F_T is uniquely determined almost everywhere with respect to the measure generated by the distribution $F_{T(j)}$ and we have

$$F_{T(j+1)}(a_{T(j+1)}) = \int_{I(a_{T(j)})} F_j^{T(j)}(a_j, x_{T(j)})F_{T(j)}(dx_{T(j)}) \tag{8-7}$$

for all $a_{T(j+1)} \in R^{T(j+1)}$, $j = 1, \ldots, n$. These formulas can be used to find the resolution. Again we may write for the sake of brevity $F_{T(j+1)} = F_j^{T(j)}F_{T(j)}$ as in the case of transition probabilities.

9.0 SIMULATION THEOREM

Let $N = \{1, 2, \ldots\}$ and $N(j) = \{k \in N: k < j\}$. Let $F_{N(n+1)}$ be a probability distribution over R^n and let $F_j^{N(j)}$ $(j = 1, \ldots, n)$ be a resolution of the distribution $F_{N(n+1)}$. If $f_1, f_2, \ldots, f_n$ are random variables over some probability space whose joint distribution is $F_{N(n+1)}$, then $F_j^{N(j)}$ is a Borel function making meaningful the conditional distribution

$$P\{f_j < a_j | f_k = a_k \text{ for all } k \in N(j)\} = F_j^{N(j)}(a_1, \ldots, a_j) \tag{9-1}$$

for all $(a_1, \ldots, a_j) \in R^j$, $j = 1, \ldots, n$.

If G is a function of j variables $a_1, \ldots, a_j$ being a probability distribution with respect to the variable a_k, denote by c_k the computerization operator acting on the k-th variable; i.e., $H = c_k(G)$ is defined by

$$H(a_1, \ldots, u_k, \ldots, a_j) = \inf\{a_k: u_k < G(a_1, \ldots, a_k, \ldots, a_j)\} \quad (9\text{-}2)$$

for all $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_j \in R$ and $u_k \in (0,1)$.

The sequence H_j $(j = 1, \ldots, n)$ of functions defined by $H_j = c_j(F_j^{N(j)})$ for $j = 1, \ldots, n$ will be called a computerization of the distribution $F_{N(n+1)}$.

9.1 LEMMA

Each function H_j for $j = 1, 2, 3, \ldots, n$ is a Borel function.

Proof. For $j = 1$, the proof is obvious since H_j is monotone. Take any $j > 1$. First, let us prove that H_j is Borel in variables $a_1, \ldots, a_{j-1}$. To this end take any number $a \in R$. Let G denote the function defined by

$$G(a_1, \ldots, a_{j-1}) = F_j^{N(j)}(a_1, \ldots, a_{j-1}, a) \quad (9\text{-}3)$$

for all $a_1, \ldots, a_{j-1} \in R$. From the equality of the sets

$$\{(a_1, \ldots, a_{j-1}): H_j(a_1, \ldots, a_{j-1}, u) < a\}$$
$$= \{(a_1, \ldots, a_{j-1}): u < G(a_1, \ldots, a_{j-1})\} \quad (9\text{-}4)$$

and the fact that the function G is Borel measurable, it follows that the function H_j is Borel measurable in the first $j-1$ variables when the jth variable u is fixed.

Since in the variable u the function H_j is monotone and right side continuous, we may conclude that H_j is Borel with respect to all its variables jointly, as we concluded in the proof of Borel measurability of a conditional distribution in theorem 8.1.

9.2 THEOREM

Let H_j $(j = 1, \ldots, n)$ be a computerization of the distribution $F_{N(n+1)}$. Define recursively the variables

$$\begin{aligned} x_1 &= H_1(u_1) \\ x_2 &= H_2(x_1, u_2) \\ x_3 &= H_3(x_1, x_2, u_3) \\ &\cdot \\ &\cdot \\ &\cdot \\ x_n &= H_n(x_1, \ldots, x_{n-1}, u_n) \end{aligned} \tag{9-5}$$

where $u_1, u_2, \ldots, u_n$ are independent random variables with uniform distribution over the open interval $(0,1)$. Then the joint probability distribution of the variables $x_1, \ldots, x_n$ coincides with the distribution $F_{N(n+1)}$.

Proof. Since each function H_j is Borel, we can prove by induction that each variable x_j as a function of the variables u_j is also Borel. Thus, x_j as functions of variables $u_1, u_2, \ldots, u_n$ are Lebesgue measurable. To find their joint distribution, we have to compute the Lebesgue measure of the set

$$D(a) = \{(u_1, \ldots, u_n) \in I^n: \ x_j < a_j \ \text{ for } \ j = 1, \ldots, n\} \tag{9-6}$$

where $I = (0,1)$. From the definition of the variables x_j and the properties of the computerization operator, we get the identity

$$D(a) = \{u \in I^n: \ u_j < F_j^{N(j)}(x_1, \ldots, x_{j-1}, a_j) \ \text{ for } \ j = 1, \ldots, n\} \tag{9-7}$$

We will prove the theorem by induction with respect to n. For $n = 1$, we have

$$D(a) = \{u \in I: \ u < F_1^{N(1)}(a)\} = (0, F_1^{N(1)}(a)) \tag{9-8}$$

and the Lebesgue measure of this set is $p(D(a)) = F_1^{N(1)}(a) = F_{N(2)}(a)$ for all $a \in R$. Assume that the theorem holds for $n = k - 1$. Notice that the set $D(a)$ can be represented in the form

$$D(a_1, \ldots, a_n)$$

$$= \{u \in I^n: \ u_1 < F_1^{N(1)}(a_1), (u_2, \ldots, u_n) \in D_{x_1}(a_2, \ldots, a_n)\} \tag{9-9}$$

where

$$D_{x_1}(a_2, \ldots, a_n)$$

$$= \{u_2, \ldots, u_n) \in I^{n-1}: \ u_j < F_j^{N(j)}(x_1, x_2, \ldots, x_{j-1}, a_j)$$

$$\text{for } j = 2, 3, \ldots, n\}$$

Notice that for almost all $x_1 = H_1(a_1)$ with respect to the measure generated by the distribution F_1 the functions $F_j^{N(j)}$ as functions of the remaining variables form a resolution of the distribution $F^1_{S(n+1)}$ in which the value of the first variable is fixed to be x_1. From Fubini's theorem, we get

$$p(D(a)) = \int_0^{F_1(a_1)} \left(\int_{D_{x_1}(a_2,\ldots,a_n)} du_2, du_3, \ldots, du_n \right) du_1$$

$$= \int_0^{F_1(a_1)} F^1_{S(n+1)}(x_1, a_2, \ldots, a_n)\, du_1 \tag{9-10}$$

From corollary 5.4, we get

$$p(D(a)) = \int_{-\infty}^{a_1} F^{1}_{S(n+1)}\,(x_1, a_2, \ldots, a_n)\, F_1(dx_1) \qquad (9\text{-}11)$$

Finally, from the properties of a conditional distribution, we get $p(D(a)) = F_{N(n+1)}\,(a_1, \ldots, a_n)$ for all $a \in R^n$. This proves the theorem.

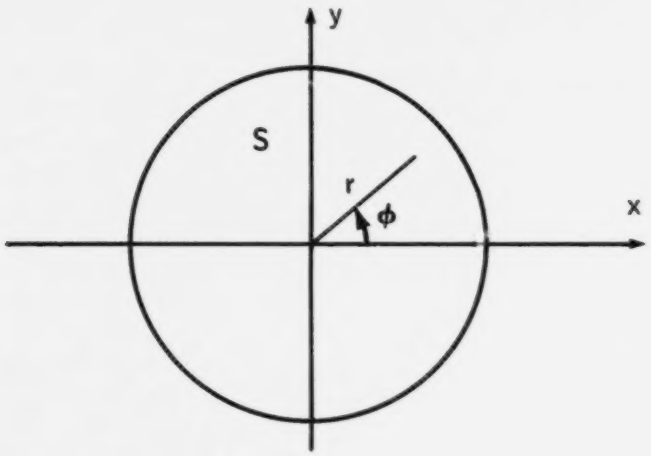
10.0 EXAMPLES OF SIMULATION

Let p be the Borel measure generated by means of Kolmogorov's construction over R^n from a joint probability distribution F of real random variables $f_1, f_2, \ldots, f_n$.

Such a sequence of random variables we shall call a random process, and the smallest closed set S in R^n whose complement has p-measure zero, we shall call the spectrum of the process. That the spectrum is well defined follows from the fact that every open set in R^n is the union of a countable family of open spheres having rational centers and rational radii. Thus, the union of all open sets of measure zero is a set of measure zero. The complement of that set is the spectrum. Clearly, to define a random process, it is sufficient to define the probability measure p over the spectrum of the process.

10.1 PROBLEM

Given is a steady flow of elementary particles through a region S in the form of a unit disk as in the following sketch.



The intensity of the current i per unit of area at a point (x, y) is given by the formula

$$i(x, y) = 1 + x \tag{10-1}$$

Consider as a random process the polar coordinates (λ, ϕ) of the point at which a particle arrives. Find a computerization of the process.

Solution. The spectrum of the process is the closed circle S. The probability density f that a particle arrives at a point with coordinates (x, y) is proportional to the intensity of the current i at the point; i.e., $f(x, y) = ci(x, y)$ for all $(x, y) \in S$. This yields the equation

$$1 = \int_S f(x, y)dxdy = c \int_S i(x, y)dxdy = c\pi \tag{10-2}$$

which yields $c = 1/\pi$.

In polar coordinates, the set S has a representation

$$S = \{(r, \phi): \ 0 < r < 1, \ 0 < \phi < 2\pi\} \tag{10-3}$$

neglecting in S several lines that have measure zero. Since $dxdy = rdrd\phi$, the probability density g in polar coordinates is given by

$$g(r, \phi) = \frac{1}{\pi} r(1 + r \cos \phi) \tag{10-4}$$

Thus, computing the distribution on the spectrum S, we get

$$F(a_1, a_2) = \frac{1}{2\pi} a_2 a_1^2 + \frac{1}{3\pi} a_1^3 \sin a_2 \tag{10-5}$$

for $0 < a_1 < 1$ and $0 < a_2 < 2\pi$. This yields the distribution $F_1^{N(1)}(a_1) = (a_1)^2$ for all $a_1 \in (0,1)$. To find the conditional distribution F_2^1, notice that the equation

$$F(a_1, a_2) = \int_{-\infty}^{a_1} F_2^1(r, a_2)\, F_1^{N(1)}(dr) \tag{10-6}$$

is equivalent to

$$\int_0^{a_1} \int_0^{a_2} g(r, \phi)\, d\phi\, dr = \int_0^{a_1} F_2^1(a_1, a_2) 2a_1 da_1 \tag{10-7}$$

Differentiating this identity with respect to a_1, we get the equation

$$\int_0^{a_2} g(a_1, \phi)\, d\phi = F_2^1(a_1, a_2) 2a_1 \tag{10-8}$$

Thus,

$$F_2^1(a_1, a_2) = \frac{1}{2a_1} \int_0^{a_2} g(a_1, \phi)\, d\phi = \frac{1}{2\pi} (a_2 + a_1 \sin a_2) \tag{10-9}$$

To find the computerization of the process (λ, ϕ), notice that from the continuity of the distribution F_1 on the interval $\langle 0,1 \rangle$ we get

$$u_1 = a_1^2, \quad \text{or} \quad a_1 = (u_1)^{1/2} \tag{10-10}$$

This yields

$$H_1(u_1) = (u_1)^{1/2} \quad \text{for all} \quad u_1 \in (0,1) \tag{10-11}$$

Similarly, by continuity of $F_2^1(a_1, a_2)$ in the second variable $a_2 \in \langle 0, 2\pi \rangle$, we get the equation

$$u_2 = \frac{1}{2\pi}(a_2 + a_1 \sin a_2) \tag{10-12}$$

This equation with respect to the variable a_2 is the Kepler equation. It can be solved either by iteration or by Newton's method.

Both methods are easily programmable on a computer. Let $a_2 = H_2(a_1, u_2)$ be the solution of the equation as a function of $a_1 \in (0,1)$ and $u_2 \in (0,1)$. The pair (H_1, H_2) represents a computerization of the process; that is, the pair of variables $\lambda = H_1(u_1)$, $\phi = H_2(\lambda, u_2)$, where u_1, u_2 are independent random variables with uniform distribution over the open unit interval $(0,1)$, will simulate the process.

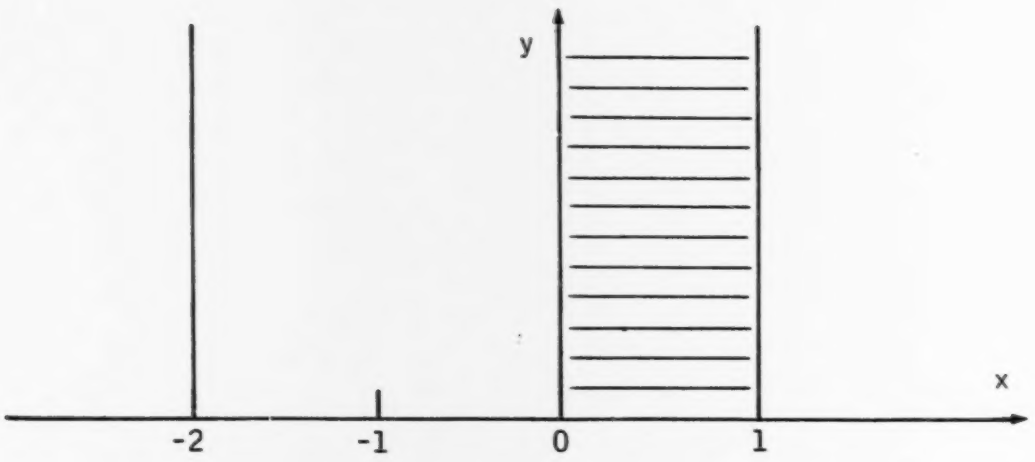
10.2 PROBLEM

Consider a chemical process generating ions. Assume that the random process consists of two variables (x, y), where x is the energy level of an ion and y is its life expectancy. Assume that the density of the probability distribution of the process is given by the formula

$$f(x, y) = h(y)e^{-2y}\,\delta(x + 2) + \frac{x}{2}(h(x) - h(x - 1))h(y)e^{-xy} \tag{10-13}$$

for all $(x, y) \in R^2$. Find a computerization of the process. (Notice that δ and h denote here, respectively, Dirac's delta function with mass centered at zero and its distribution function over R.)

Solution. Notice that the spectrum of the process is given by the following diagram.



It consists of an infinite ray at the position $x = -2$ and an infinite strip above the x-axis such that $0 \leq x \leq 1$.

Computing the distribution function F, we get

$$\left.\begin{aligned}
&F(a_1, a_2) = 0 && \text{if } a_2 \leq 0 \text{ and } a_1 \in R \\
&F(a_1, a_2) = 0 && \text{if } a_2 > 0 \text{ and } a_1 \leq -2 \\
&F(a_1, a_2) = (1 - e^{-2a_2})/2 && \text{if } a_2 > 0 \text{ and } -2 < a_1 \leq 0 \\
&F(a_1, a_2) = (1/2)(1 - e^{-2a_2}) + (1/2)(a_1 + a_2^{-1}(e^{-a_1 a_2} - 1)) \\
&\qquad\qquad \text{if } a_2 > 0 \text{ and } 0 < a_1 \leq 1 \\
&F(a_1, a_2) = (1/2)(1 - e^{-2a_2}) + (1/2)(1 + a_2^{-1}(e^{-a_2} - 1)) \\
&\qquad\qquad \text{if } a_2 > 0 \text{ and } a_1 > 1
\end{aligned}\right\} \quad (10\text{-}14)$$

This set of equations yields the formulas

$$\left.\begin{aligned} F_1^{N(1)}(a_1, a_2) &= 0 && \text{if } a_1 \leq -2 \\ F_1^{N(1)}(a_1, a_2) &= \frac{1}{2} && \text{if } -2 < a_1 \leq 0 \\ F_1^{N(1)}(a_1, a_2) &= \frac{1}{2}(1 + a_1) && \text{if } 0 < a_1 \leq 1 \\ F_1^{N(1)}(a_1, a_2) &= 1 && \text{if } a_1 > 1 \end{aligned}\right\} \qquad (10\text{-}15)$$

The spectrum of the variable x consists of the point $x = -2$ and the interval $\langle 0,1\rangle$. At the point $x = -2$, the measure generated by the distribution $F_1^{N(1)}$ has mass 1/2, and on the interval $\langle 0,1\rangle$ it has a linear mass density equal to 1/2. Using these properties, we get for the conditional distribution F_2^1 the values

$$\left.\begin{aligned} F_2^1(a_1, a_2) &= 1 - e^{-2a_2} && \text{if } a_1 = -2 \\ F_2^1(a_1, a_2) &= 1 - e^{-a_1 a_2} && \text{if } 0 < a_1 < 1 \end{aligned}\right\} \qquad (10\text{-}16)$$

Since outside the spectrum of the variable x the function F_2^1 may be defined arbitrarily, let us set

$$\left.\begin{aligned} F_2^1(a_1, a_2) &= 1 - e^{-2a_2} && \text{if } a_1 \leq 0 \\ F_2^1(a_1, a_2) &= 1 - e^{-a_1 a_2} && \text{if } a_1 > 0 \end{aligned}\right\} \qquad (10\text{-}17)$$

From the graph of the function $F_1^{N(1)}$, we get the formula

$$\left.\begin{aligned} H_1(u_1) &= -2 && \text{if } 0 < u_1 < 1/2 \\ H_1(u_1) &= 2u_1 - 1 && \text{if } 1/2 \leq u_1 < 1 \end{aligned}\right\} \qquad (10\text{-}18)$$

Since the function $F_2^1(a_1, a_2)$ is continuous in the variable a_2 for a fixed value of the variable a_1 and invertible on the interval $<0,\infty)$, we get

$$\left.\begin{aligned} H_2(a_1, u_2) &= -(1/2)\log(1 - u_2) && \text{if } a_1 \leq 0 \\ H_2(a_1, u_2) &= -(1/a_1)\log(1 - u_2) && \text{if } a_1 > 0 \end{aligned}\right\} \qquad (10\text{-}19)$$

The pair of functions (H_1, H_2) represents a computerization of the process x, y. In the above example, we considered for the sake of simplicity a process having only one spectral line and one continuous areal component. The method used here can be easily extended to the case where the spectrum consists of a sequence of spectral lines and several two-dimensional components.

11.0 EXPECTED VALUE OF A FUNCTION OF A PROCESS

Many applications require computation of some parameters of a process, such as covariance matrix, moments, and characteristic function. These computations require one to find the expectation

$$E(f(x_1, x_2, \ldots, x_k)) \qquad (11\text{-}1)$$

where f is a sufficiently regular Borel function defined on the range (spectrum) of the process $x_1, x_2, \ldots, x_k$.

Computer simulation allows one to find the approximate values of the expectation and to establish probabilistic bounds on the error of the expected value. This can be done by involving the central limit theorem if the function f has a finite second moment. Treating the value

$$y = f(x_1, x_2, \ldots, x_k) \qquad (11\text{-}2)$$

as a random variable, one may find by simulation of the process sufficiently large stochastically independent samples of the variable y. The mean of the sample will approximate the expectation E(y). Since the mean for large samples has approximately normal distribution, from the sample of the variable y one can easily estimate the variance of the mean, and thus get an idea of the accuracy of the estimate of the expected value.

12.0 INTEGRAL FORMULA FOR EXPECTATION

Let F be a probability distribution over R^n and $H_1, H_2, \ldots, H_n$ its computerization. Define a map G from the cube I^n, where $I = (0,1)$, into R^n by the formula $x = G(u)$, where

$$\begin{aligned} x_1 &= H_1(u_1) \\ x_2 &= H_2(x_1, u_2) \\ x_3 &= H_3(x_1, x_2, u_3) \\ &\cdot \\ &\cdot \\ &\cdot \\ x_n &= H_n(x_1, \ldots, x_{n-1}, u_n) \end{aligned} \qquad (12\text{-}1)$$

for all $u \in I^n$. Let p be the Borel measure corresponding to the probability distribution F. Let m be the classical Lebesgue measure over the cube I^n.

12.1 THEOREM

The map K defined by $K(f) = foG$ establishes linear isometric imbedding of the Lebesgue space $L(p, R)$ of summable functions into the Lebesgue space $L(m, R)$.

The proof of the theorem is similar to the corresponding proof for one variable presented in section 5.

Corollary. If the right-hand side integral in the following formula exists, then so does the other and they are equal:

$$\int_{R^n} f(x)dF = \int_{I^n} f(G(u))du \qquad (12\text{-}2)$$

Definitions of the integral are similar to those in section 5.

13.0 CONCLUSION

The principal result of this paper is the proof of the existence of a recursive algorithm by means of which one can simulate on the computer any finite sequence $x_1, x_2, \ldots, x_k$ of random variables whose joint distribution F is known. These variables may be dependent and their joint spectrum may have continuous and discrete components.

This result should be useful in applications requiring the Monte Carlo method, in particular in problems of quantum chemistry, nuclear and plasma physics, economics, and stochastic control systems.

A word of caution is appropriate here. Since most computer languages use words of a fixed number of bits to represent numbers, the set of numbers available in such languages is finite. Even if one used some set of computable real numbers, say all rationals, as one could define, for example, by means of the PL-language of Brainerd-Landweber (ref. 7), the set of all numbers available on the computer would be at most countable and thus of Lebesgue measure zero. Hence, it is always possible to find a pathological example of a distribution F whose computerization H cannot be simulated by a computer. However, in most applications the resulting computerization H consists of functions that can be well approximated by means of piecewise continuous functions whose computational complexity is not too great.

14.0 REFERENCES

1. Bogdan, V. M.: A New Approach to the Theory of Probability Via Algebraic Categories. Lecture Notes in Mathematics, vol. 541, Springer-Verlag, 1975, pp. 345-367.

2. Bogdanowicz, W. M.: A Generalization of the Lebesgue-Bochner Integral and a New Approach to the Theory of Integration. Proc. Nat. Acad. Sci. U.S., vol. 53 (1965), pp. 492-498.

3. Bogdanowicz, W. M.: An Approach to the Theory of Lebesgue-Bochner Measurable Functions and to the Theory of Measure. Mathematische Annalen, vol. 164 (1966), pp. 251-269.

4. Bogdanowicz, W. M.: An Approach to the Theory of Integration and Theory of Lebesgue-Bochner Measurable Functions on Locally Compact Spaces. Mathematische Annalen, vol. 171 (1967), pp. 219-238.

5. Bogdanowicz, W. M.: Relations Between the Lebesgue Integral Generated by a Measure and the Integral Generated by a Volume. Commentationes Mathematicae, vol. 12 (1969), pp. 277-299.

6. Bogdanowicz, W. M.; and Kritt, B.: Radon-Nikodym Differentiation of One Vector-Valued Volume with Respect to Another. Bulletin de l'Academie Polonaise des Science, vol. 15 (1967), pp. 479-485.

7. Brainerd, W. S.; and Landweber, L. H.: Theory of Computation. John Wiley and Sons, 1974.

8. Halmos, P. R.: Measure Theory. D. Van Nostrand, 1950.

9. Kolmogorov, A. N.: Foundations of the Theory of Probability. Chelsea, 1950.

10. Neveu, T.: Mathematical Foundations of the Calculus of Probability. Holden-Day, 1965.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TP-1859		

4. Title and Subtitle	5. Report Date
COMPUTER SIMULATION OF RANDOM VARIABLES AND VECTORS WITH ARBITRARY PROBABILITY DISTRIBUTION LAWS	May 1981
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Victor M. Bogdan	S-504
	10. Work Unit No. 910-44-LH-00-72

9. Performing Organization Name and Address	11. Contract or Grant No.
Lyndon B. Johnson Space Center Houston, Texas 77058	
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Technical Paper
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

This research was done while the author was a National Research Council Senior Resident Research Associate during a sabbatical leave from the Catholic University of America in 1979.

16. Abstract

Assume that there is given an arbitrary n-dimensional probability distribution F. In this paper, a recursive construction is found for a sequence of functions $x_1 = f_1(U_1, \ldots, U_n), \ldots, x_n = f_n(U_1, \ldots, U_n)$ such that if $U_1, \ldots, U_n$ are independent random variables having uniform distribution over the open interval (0,1), then the joint distribution of the variables $x_1, \ldots, x_n$ coincides with the distribution F. Since uniform independent random variables can be well simulated by means of a computer, this result allows one to simulate arbitrary n-random variables if their joint probability distribution is known.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Computerized simulation Expectancy hypothesis Expectation Monte Carlo method Probability distribution functions Transition probabilities	Unclassified - Unlimited Subject Category 65

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	43	A03

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1981